Weighted Model Counting with Conditional Weights for Bayesian Networks

Paulius Dilkas, Vaishak Belle

University of Edinburgh, Edinburgh, UK
p.dilkas@sms.ed.ac.uk, vaishak@ed.ac.uk

Abstract

Weighted model counting (WMC) has emerged as the **unifying inference mechanism** across many (probabilistic) domains. Encoding an inference problem as an instance of WMC typically **necessitates adding extra literals and clauses**. This is partly so because the predominant definition of WMC assigns weights to models based on weights on literals, and this severely restricts what probability distributions can be represented. We develop a **measure-theoretic perspective** on WMC and propose a way to encode conditional weights on literals analogously to conditional probabilities. This representation can be as succinct as standard WMC with weights on literals but can also expand as needed to represent probability distributions with less structure. To demonstrate the performance benefits of conditional weights over the addition of extra literals, we develop a **new WMC encoding for Bayesian networks** `cw` and adapt a state-of-the-art WMC algorithm ADDMC to the new format. Our experiments show that the new encoding significantly improves the performance of the algorithm on most benchmark instances.

WMC as a Measure on a Boolean Algebra

Traditionally, WMC is seen as a **generalisation of propositional model counting**. Here we provide a measure-theoretic definition, the main benefit of which is that it allows us to **define weights more flexibly** and results in **significantly smaller encodings**. While measures on Boolean algebras have been studied before, other definitions are new. For any set U, 2^{2^U} (i.e., the powerset of the powerset of U) is the Boolean algebra that directly corresponds to **all possible propositional formulas** constructable from a and b. A **measure** on 2^{2^U} is a function $\mu\colon 2^{2^U} \to \mathbb{R}_{\geq 0}$ such that

- $\mu(\bot) = 0$, and
- $\mu(a \lor b) = \mu(a) + \mu(b)$ for all $a, b \in 2^{2^U}$ whenever $a \land b = \bot$.

A **weight function** is a function $\nu\colon 2^U \to \mathbb{R}_{\geq 0}$. A weight function is **factored** if $\nu = \prod_{x \in U} \nu_x$ for some functions $\nu_x\colon 2^{\{x\}} \to \mathbb{R}_{\geq 0}$, $x \in U$. We say that a weight function $\nu\colon 2^U \to \mathbb{R}_{\geq 0}$ **induces** a measure $\mu_\nu\colon 2^{2^U} \to \mathbb{R}_{\geq 0}$ if

$$\mu_\nu(x) = \sum_{\{u\} \leq x} \nu(u).$$

Finally, a measure $\mu\colon 2^{2^U} \to \mathbb{R}_{\geq 0}$ is **factorable** if there exists a factored weight function $\nu\colon 2^U \to \mathbb{R}_{\geq 0}$ that induces μ. In this formulation, WMC corresponds to **the process of calculating the value of** $\mu_\nu(x)$ for some $x \in 2^{2^U}$ with a given definition of ν. Moreover, **classical WMC is only able to evaluate factorable measures.**

Relation to the Classical (Logic-Based) View of WMC

The classical definition of WMC relies on a **weight function over the literals** of a propositional theory. Let $\mathcal{L}$ be a propositional logic with two atoms a and b and $w\colon \{a, b, \neg a, \neg b\} \to \mathbb{R}_{\geq 0}$ a weight function defined as $w(a) = 0.3$, $w(\neg a) = 0.7$, $w(b) = 0.2$, $w(\neg b) = 0.8$. Furthermore, let Δ be a theory in $\mathcal{L}$ with a sole axiom a. Then Δ has two models: $\{a, b\}$ and $\{a, \neg b\}$ and its WMC is

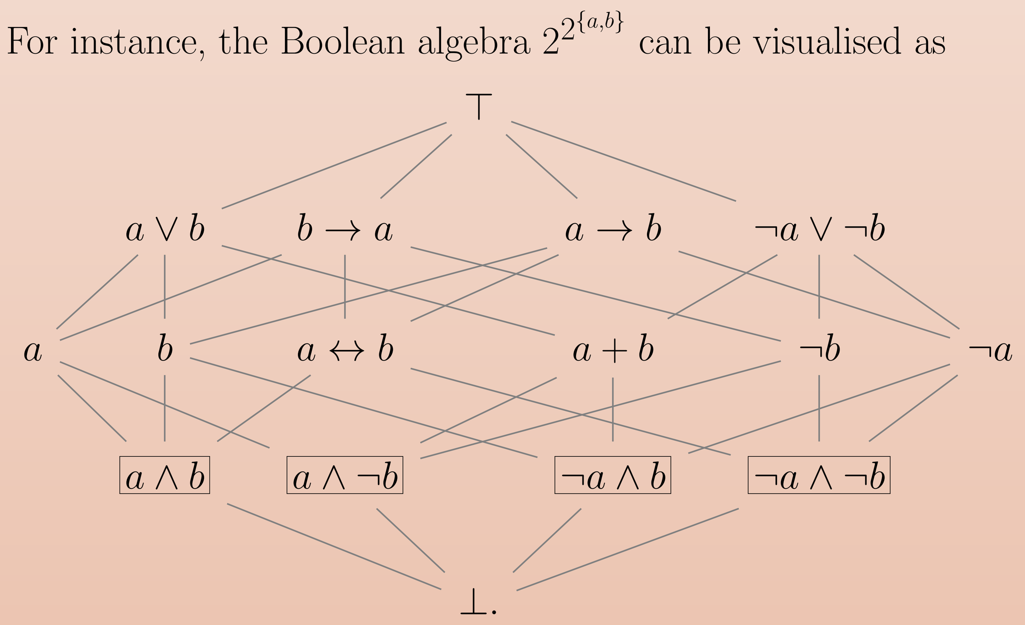
$$\mathrm{WMC}(\Delta) = \sum_{\omega \models \Delta} \prod_{\omega \models l} w(l) = w(a)w(b) + w(a)w(\neg b) = 0.3, \tag{1}$$

i.e., we **sum over all models** entailed by the theory, and the weight of each model is **the product of the weights of all literals in it**. Alternatively, we can define $\nu_a\colon 2^{\{a\}} \to \mathbb{R}_{\geq 0}$ as $\nu_a(\{a\}) = 0.3$, $\nu_a(\emptyset) = 0.7$ and $\nu_b\colon 2^{\{b\}} \to \mathbb{R}_{\geq 0}$ as $\nu_b(\{b\}) = 0.2$, $\nu_b(\emptyset) = 0.8$. Let μ be the measure on 2^{2^U} induced by $\nu = \nu_a \cdot \nu_b$. Then, equivalently to Eq. (1), we can write

$$\begin{aligned}\mu(a) &= \nu(\{a, b\}) + \nu(\{a\}) \\ &= \nu_a(\{a\})\nu_b(\{b\}) + \nu_a(\{a\})\nu_b(\emptyset) = 0.3.\end{aligned}$$

Thus, one can equivalently think of WMC as summing over **models of a theory** or over **atoms below an element** of a Boolean algebra.

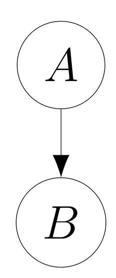
For instance, the Boolean algebra $2^{2^{\{a,b\}}}$ can be visualised as



Boxed elements are the **atoms** of the algebra.

Example: Encoding Bayesian Networks

A typical WMC encoding for the pictured Bayesian network needs **8 variables and 17 clauses**. However, the `cw` weight function only needs **two variables**. Let $U = \{\lambda_{A=1}, \lambda_{B=1}\}$ be the set of variables. Using some simple syntax for manipulating weight functions (mostly defined pointwise), we can define the weight function $\nu\colon 2^U \to \mathbb{R}_{\geq 0}$ for this network as $\nu \coloneqq \nu_A \cdot \nu_B$, where



a	$\Pr(A = a)$
1	0.5
0	0.5

a	b	$\Pr(B = b \mid A = a)$
1	1	0.6
1	0	0.4
0	1	0.1
0	0	0.9

Fig. 1: A Bayesian network with its conditional probability tables

$$\nu_A = 0.5,$$

and

$$\nu_B = 0.6[\lambda_{B=1}] \cdot [\lambda_{A=1}] + 0.4[\neg\lambda_{B=1}] \cdot [\lambda_{A=1}] + 0.1[\lambda_{B=1}] \cdot [\neg\lambda_{A=1}] + 0.9[\neg\lambda_{B=1}] \cdot [\neg\lambda_{A=1}].$$

Experimental Results

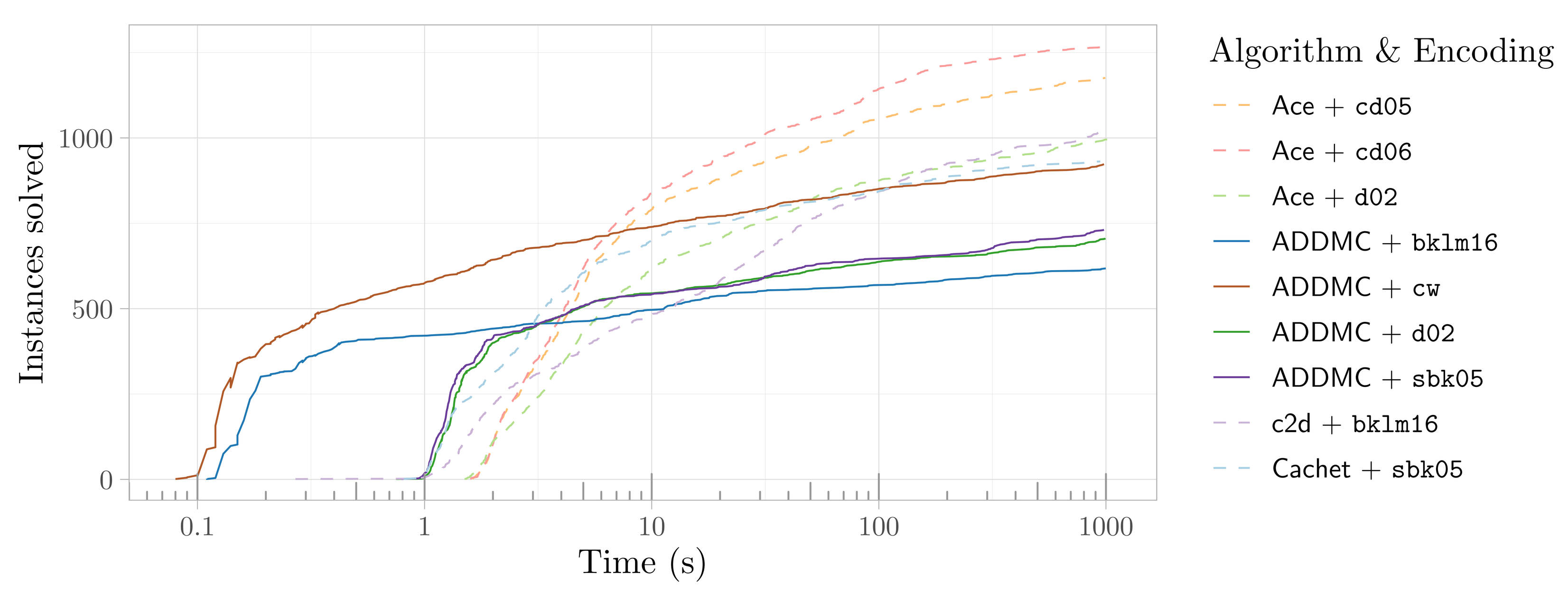


Fig. 2: Cumulative numbers of instances solved by combinations of algorithms and encodings over time

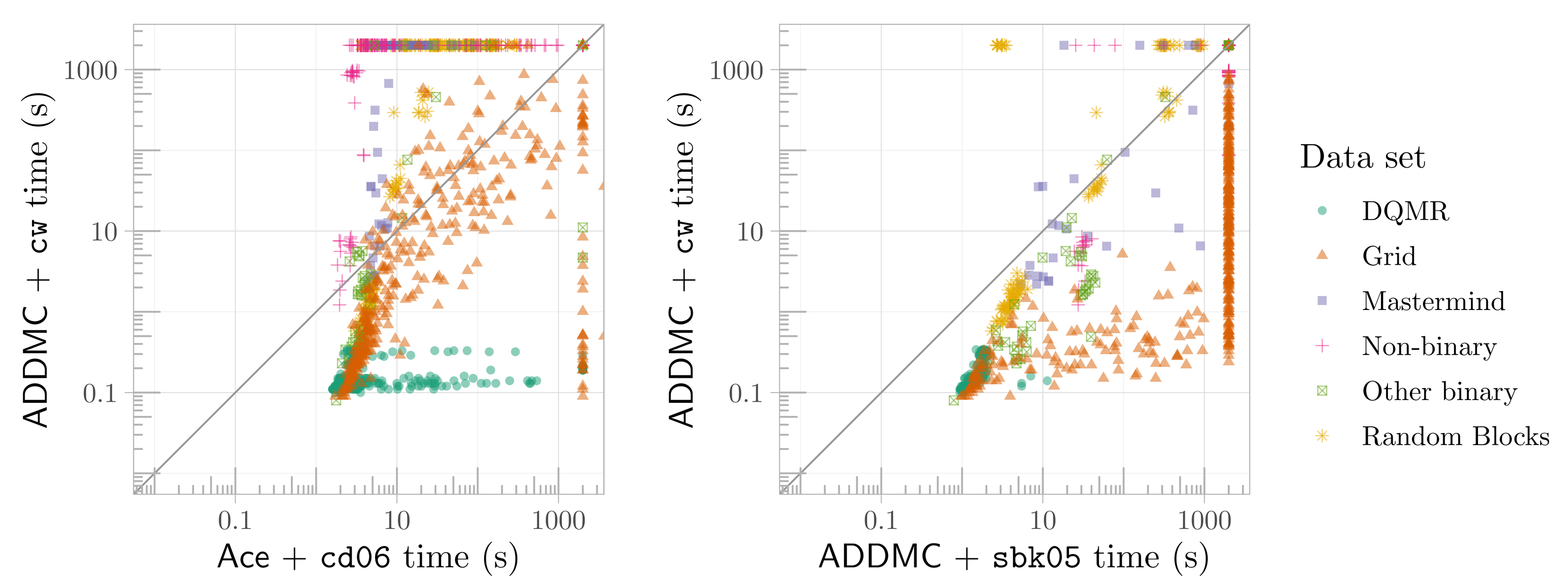


Fig. 3: An instance-by-instance comparison between ADDMC + `cw` and the best overall combination of algorithm and encoding (Ace + `cd06`, on the left) as well as the second-best encoding for ADDMC (`sbk05`, on the right)

We compare all six WMC encodings for Bayesian networks when run with both ADDMC and the WMC algorithms used in the original papers. The cumulative plot shows that ADDMC significantly **underperforms** when combined with any of the **previous encodings**, but our encoding `cw` significantly **improves** the performance of ADDMC, making ADDMC + `cw` comparable to many other combinations. The scatter plot on the left-hand side adds to this by showing that `cw` is **particularly promising on certain data sets** such as DQMR and Grid networks. The scatter plot on the right-hand side shows that `cw` is better than `sbk05` (i.e., the second-best encoding for ADDMC) on the **majority of instances**. Seeing how, e.g., DQMR instances are trivial for ADDMC + `cw` but hard for Ace + `cd06`, and vice versa for Mastermind instances, we conclude that the best-performing algorithm-encoding combination **depends** significantly **on** (as-of-yet unknown) **properties of the Bayesian networks**.

See the Paper for...

- Theoretical results that establish both the limits and the capabilities of classical WMC.
- Details of the `cw` encoding for Bayesian networks.
- A more detailed comparison of Bayesian network encodings, including a parameterized asymptotic analysis of their growth.

THE UNIVERSITY of EDINBURGH informatics

EDINBURGH CENTRE FOR ROBOTICS

UKRI Engineering and Physical Sciences Research Council